

Klasifikasi Penyakit Menular Dengan Algoritma *Random Forest* Dan SVM

Ardiansyah Dwi Prasetyo^{1*}, Dhiya Muthi Zafarani², Muhammad Ammar³

^{1,3}Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspipetek No. 46,
Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: ^{1*}ardiansyah134765@gmail.com, ²dhiyamuthiz@gmail.com,

³muhammadammarr23@gmail.com

(* : coressponding author)

Abstrak - Penyakit menular adalah masalah kesehatan global yang memerlukan deteksi dini dan akurat untuk mencegah penyebaran lebih lanjut. Penelitian ini bertujuan untuk membandingkan kinerja dari algoritma pembelajaran mesin ini, yaitu Random Forest (RF) dan Support Vector Machine (SVM), dalam klasifikasi penyakit menular berdasarkan data kesehatan. Dataset yang digunakan mencakup berbagai fitur medis seperti gejala klinis, riwayat kontak, dan hasil pemeriksaan laboratorium. Kedua algoritma diuji dengan teknik cross-validation untuk mengukur akurasi. Hasil dari penelitian ini menunjukkan bahwa algoritma RF dan SVM keduanya mampu mencapai akurasi yang baik dalam mengidentifikasi penyakit menular. RF menunjukkan kinerja yang unggul dalam hal stabilitas dan kemampuan menangani data dengan banyak fitur, sementara SVM menunjukkan performa optimal pada data dengan pemisahan yang jelas antar kelas. Kesimpulannya, baik RF maupun SVM memiliki potensi sebagai alat bantu diagnosis penyakit menular, dengan pemilihan algoritma yang bergantung pada karakteristik data yang tersedia.

Kata Kunci: Penyakit Menular, Klasifikasi, Random Forest, Support Vector Machine, Pembelajaran Mesin, Akurasi.

Abstract - Infectious diseases are a global health problem that require early and accurate detection to prevent further spread. This research aims to compare the performance of these machine learning algorithms, namely Random Forest (RF) and Support Vector Machine (SVM), in the classification of infectious diseases based on health data. The dataset used includes various medical features such as clinical symptoms, contact history, and laboratory examination results. Both algorithms were tested using cross-validation techniques to measure accuracy. The results of this study show that the RF and SVM algorithms are both able to achieve good accuracy in identifying infectious diseases. RF shows superior performance in terms of stability and ability to handle data with many features, while SVM shows optimal performance on data with clear separation between classes. In conclusion, both RF and SVM have potential as tools to diagnose infectious diseases, with the choice of algorithm depending on the characteristics of the available data.

Keywords. Infectious Disease, Classification, Random Forest, Support Vector Machine, Machine Learning, Accuracy

1. PENDAHULUAN

Penyakit menular adalah salah satu tantangan utama di bidang kesehatan yang terus berkembang, dengan potensi penyebaran yang cepat dan berdampak besar pada populasi. Penyakit ini dapat disebabkan oleh berbagai agen infeksi seperti virus, bakteri, parasit, atau sepora, dan dapat menyebar melalui berbagai cara seperti udara, kontak langsung, atau media perantara lainnya. Peningkatan jumlah kasus penyakit menular, terutama yang bersifat epidemik atau pandemik, seperti COVID-19, Ebola, dan influenza, menyoroti pentingnya deteksi dini dan klasifikasi penyakit yang tepat untuk mencegah penyebaran lebih lanjut dan untuk mendukung tindakan medis yang efektif.

Kemajuan dalam teknologi informasi dan kecerdasan buatan, khususnya pembelajaran mesin, memberikan peluang untuk mengembangkan sistem klasifikasi otomatis dalam mendeteksi penyakit menular. Algoritma pembelajaran mesin mampu menganalisis data dalam jumlah besar dan mengidentifikasi pola yang sulit dikenali oleh manusia. Hal ini membuka peluang untuk memanfaatkan teknik pembelajaran mesin dalam membantu tenaga medis membuat keputusan yang lebih cepat dan akurat dalam mendiagnosis penyakit menular.

Dalam penelitian, algoritma Random Forest (RF) dan Support Vector Machine (SVM) digunakan untuk membangun model klasifikasi penyakit menular. Random Forest adalah algoritma berbasis ensemble yang terdiri dari banyak pohon keputusan, yang memberikan keunggulan dalam

hal akurasi dan ketahanan terhadap overfitting pada dataset yang kompleks. Di sisi lain, Support Vector Machine dikenal efektif dalam menangani data berdimensi tinggi dan mampu memisahkan kelas-kelas dalam data dengan margin yang optimal.

Penelitian ini tujuannya untuk mengkaji performa kedua algoritma dalam mengklasifikasikan penyakit menular dengan akurasi tinggi. Dengan menggunakan data gejala, riwayat medis, dan faktor risiko lainnya, model yang dihasilkan diharapkan dapat mendukung tenaga medis dalam mendiagnosis penyakit menular secara lebih efektif dan efisien.

2. METODE PENELITIAN

Pada Metode Penelitian ini dilakukan dengan mengembangkan model klasifikasi penyakit menular menggunakan algoritma Random Forest (RF) dan Support Vector Machine (SVM). Data yang digunakan dalam penelitian ini dikumpulkan dari dataset medis yang mencakup informasi seperti gejala klinis, riwayat kesehatan, faktor risiko, dan hasil diagnosis. Dataset yang digunakan memungkinkan pengklasifikasian beberapa jenis infeksi umum, dengan tetap memperhatikan aspek privasi dan keamanan data pasien.

2.1 Study Literatur Review (SLR)

Study Literatur Review (SLR) ini bertujuan untuk memberikan gambaran tentang penerapan kedua algoritma ini dalam klasifikasi penyakit menular, membandingkan keunggulan dan keterbatasan masing-masing model, serta mengidentifikasi tantangan yang dihadapi dalam penerapan teknik ini.

2.2 Pengumpulan Data

Pengumpulan data adalah langkah awal yang sangat penting dalam penerapan model pembelajaran mesin, baik itu Random Forest (RF) maupun Support Vector Machine (SVM), dalam klasifikasi penyakit menular. Data yang dikumpulkan harus representatif, akurat, dan relevan agar model yang dibangun dapat memberikan prediksi yang baik. Dalam konteks penyakit menular, data yang digunakan biasanya mencakup berbagai fitur medis yang dapat mempengaruhi diagnosis atau deteksi penyakit.

2.3 Preprocessing Data

Langkah pertama dalam preprocessing adalah pengumpulan dan pembersihan data. Ini mencakup beberapa kegiatan yang bertujuan untuk memastikan bahwa data yang akan digunakan berkualitas tinggi.

2.4 Penerapan Model Random Forest Dan Support Vector Machine (SVM)

Penerapan algoritma Random Forest (RF) dan Support Vector Machine (SVM) dalam klasifikasi penyakit menular merupakan salah satu pendekatan yang efektif dalam memanfaatkan teknologi pembelajaran mesin untuk mendeteksi penyakit secara cepat dan akurat. Kedua model ini memiliki karakteristik yang berbeda dalam menangani masalah klasifikasi, terutama pada data kesehatan yang sering kali memiliki kompleksitas tinggi dan ketidakseimbangan antar kelas.

2.5 Evaluasi Dan Perbandingan Metode

Random Forest menggunakan kumpulan decision tree untuk membuat prediksi. Pendekatan ini memungkinkan model bekerja dengan baik pada dataset yang besar dan sering mengandung noise atau data yang tidak konsisten. Selain itu, Random Forest bisa memberikan informasi tentang fitur-fitur yang berperan penting dalam klasifikasi, yang sangat berguna untuk analisis faktor risiko dalam penyakit menular. Namun, karena terdiri dari banyak pohon keputusan, Random Forest membutuhkan lebih banyak waktu dan sumber daya komputasi.

SVM bekerja dengan cara menemukan hyperplane yang optimal untuk memisahkan kelas-kelas dalam dataset. SVM umumnya cocok untuk dataset yang lebih kecil namun kompleks, karena cenderung memberikan hasil yang akurat dan tahan terhadap overfitting dengan menggunakan margin yang optimal. Pemilihan parameter kernel yang tepat sangat berpengaruh pada performa SVM, dan algoritma ini lebih sensitif terhadap data outlier.

2. ANALISA DAN PEMBAHASAN

Hasil studi literature review sebagai berikut:

Tabel 1. Jenis Jenis *Database*

No	Nama Peneliti Dan Tahun	Metode penelitian	Tujuan Penelitian	Manfaat
1	Fauzan Azimah, Kiky Risky Nova wardani , 2021	Random Forest Dan Support Vector Machine	untuk mengetahui tingkat risiko terjangkit virus COVID-19 berdasarkan gejala yang ditimbulkan, dengan menggunakan beberapa algoritma klasifikasi untuk memprediksi kemungkinan seseorang terinfeksi virus berdasarkan data yang tersedia.	Manfaat dari penelitian ini antara lain adalah memberikan sistem klasifikasi yang dapat membantu memprediksi risiko seseorang terinfeksi COVID-19 berdasarkan gejala yang muncul, yang pada gilirannya dapat mempermudah proses penilaian risiko secara cepat dan akurat. Penelitian ini juga menunjukkan penerapan algoritma Random Forest dan Support Vector Machine (SVM) di bidang kesehatan, khususnya dalam memprediksi infeksi virus, yang dapat digunakan untuk pengambilan keputusan medis yang lebih efektif.
2	Chea Zahrah Vaganza Junus, Tarno, Puspita Kartikasari , 2022	Random Forest Dan Klasifikasi Linear	Penelitian ini bertujuan untuk mengetahui metode mana yang memiliki performa lebih baik dalam mengklasifikasikan individu dengan risiko Diabetes Melitus berdasarkan faktor-faktor yang relevan.	Manfaat dari penelitian ini antara lain meningkatkan akurasi deteksi dini Diabetes Melitus, memungkinkan intervensi lebih cepat dan efektif bagi individu yang berisiko tinggi. Penelitian ini juga memperluas penerapan <i>machine learning</i> di bidang kesehatan, khususnya untuk memprediksi penyakit tidak menular secara akurat. masyarakat dan tenaga medis. Dengan deteksi dini yang lebih akurat, risiko komplikasi serius seperti kebutaan, penyakit jantung, dan gagal ginjal pada penderita Diabetes Melitus dapat dikurangi, sehingga kualitas hidup penderita meningkat. Penelitian ini juga memberikan edukasi mengenai pentingnya deteksi dini dan pemahaman faktor risiko, sehingga dapat lebih waspada dan mengambil langkah pencegahan yang tepat.
3	Cindy, Tiffany Sabatini, Vincent Itan , 2024	Support Vector Machine (SVM)	Tujuan penelitian ini adalah untuk menganalisis gejala monkeypox menggunakan metode Support Vector Machine (SVM) guna meningkatkan	Manfaat dari penelitian ini adalah untuk mendukung upaya pencegahan dan penanganan kasus monkeypox agar tidak mengalami peningkatan. Dengan menggunakan SVM, penelitian ini memberikan solusi yang lebih akurat dalam mendiagnosa monkeypox dibandingkan metode konvensional. Pendekatan ini juga

			<p>akurasi klasifikasi dalam mendiagnosa monkeypox. Selain itu, penelitian ini juga bertujuan mengatasi ketidakseimbangan kelas pada dataset gejala monkeypox melalui kombinasi metode Random Oversampling (ROS) dan Random Undersampling (RUS) sehingga dapat memprediksi pola gejala monkeypox secara lebih efektif.</p>	<p>memungkinkan klasifikasi gejala monkeypox yang lebih baik, sehingga hasilnya dapat membantu memahami penyebab dan menentukan penanganan yang tepat.</p>
--	--	--	--	--

4. KESIMPULAN

Berdasarkan tinjauan literatur algoritma Random Forest (RF) dan Support Vector Machine (SVM) digunakan untuk membangun model klasifikasi penyakit menular. Random Forest adalah algoritma berbasis ensemble yang terdiri dari banyak pohon keputusan, yang memberikan keunggulan dalam hal akurasi dan ketahanan terhadap overfitting pada dataset yang kompleks.

Penelitian ini membandingkan performa algoritma Random Forest (RF) dan Support Vector Machine (SVM) dalam klasifikasi penyakit menular. Hasil analisis menunjukkan bahwa keduanya mampu mengklasifikasikan penyakit menular dengan akurasi yang baik. Random Forest unggul dalam menangani dataset yang kompleks dan cenderung stabil meskipun terdapat banyak fitur dalam data. Di sisi lain, Support Vector Machine menunjukkan performa yang optimal pada data yang memiliki pemisahan yang jelas antar kelas. Oleh karena itu, kedua algoritma memiliki potensi untuk digunakan dalam mendukung diagnosis penyakit menular, dengan pemilihan algoritma yang disesuaikan berdasarkan karakteristik data yang tersedia.

REFERENCES

- Mulajati, Muhammad. (2017). IMPLEMENTASI TEKNIK WEB SCRAPING DAN KLASIFIKASI SENTIMEN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER DAN ASOSIASI TEKS (*Studi Kasus: Data Ulasan Penumpang Maskapai Penerbangan Garuda Indonesia Pada Situs TripAdvisor*)
- E. Fitri, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *J. Transform.*, vol. 18, no. 1, p. 71, 2020, doi: 10.26623/transformatika.v18i1.2317.
- Muiz Suyaana, A., Juni Aditya, B., Fayza Pramestia, P., Cahya Maharani, V., & Rosyani, P. (2024). Klasifikasi Penyakit Menular Dengan Algoritma Machine Learning Berbasis SVM. *JRIIN :Jurnal Riset Informatika Dan Inovasi*, 2(7), 1207–1212. Retrieved from
- P. Elisa and A. R. Isnain, "COMPARISON OF RANDOM FOREST, SUPPORT VECTOR MACHINE AND NAIVE BAYES ALGORITHMS TO ANALYZE SENTIMENT TOWARDS MENTAL HEALTH STIGMA", *J. Tek. Inform. (JUTIF)*, vol. 5, no. 1, pp. 321-329, Feb. 2024.
- Purbolaksono, M. D., Irvan Tantowi, M., Imam Hidayat, A., & Adiwijaya, A. (2021). Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(2), 393-399.